Spatial mixing and the connective constant: Optimal bounds

Alistair Sinclair* Piyush Srivastava† Daniel Štefankovič‡ Yitong Yin§

Abstract

We study the problem of deterministic approximate counting of matchings and independent sets in graphs of bounded *connective constant*. More generally, we consider the problem of evaluating the partition functions of the *monomer-dimer model* (which is defined as a weighted sum over all matchings where each matching is given a weight $\gamma^{|V|-2|M|}$ in terms of a fixed parameter γ called the *monomer activity*) and the *hard core model* (which is defined as a weighted sum over all independent sets where an independent set I is given a weight $\lambda^{|I|}$ in terms of a fixed parameter λ called the *vertex activity*). The *connective constant* is a natural measure of the average degree of a graph which has been studied extensively in combinatorics and mathematical physics, and can be bounded by a constant even for certain unbounded degree graphs such as those sampled from the sparse Erdős–Rényi model $\mathcal{G}(n, d/n)$.

Our main technical contribution is to prove the best possible rates of decay of correlations in the natural probability distributions induced by both the hard core model and the monomer-dimer model in graphs with a given bound on the connective constant. These results on decay of correlations are obtained using a new framework based on the so-called *message* approach that has been extensively used recently to prove such results for bounded degree graphs. We then use these optimal decay of correlations results to obtain FPTASs for the two problems on graphs of bounded connective constant.

In particular, for the monomer-dimer model, we give a deterministic FPTAS for the partition function on all graphs of bounded connective constant for any given value of the monomer activity. The best previously known deterministic algorithm was due to Bayati, Gamarnik, Katz, Nair and Tetali [STOC 2007], and gave the same runtime guarantees as our results but only for the case of bounded degree graphs. For the hard core model, we give an FPTAS for graphs of connective constant Δ whenever the vertex activity $\lambda < \lambda_c(\Delta)$, where $\lambda_c(\Delta) := \frac{\Delta^\Delta}{(\Delta-1)^{\Delta+1}}$; this result is optimal in the sense that an FPTAS for any $\lambda > \lambda_c(\Delta)$ would imply that NP=RP [Sly, FOCS 2010]. The previous best known result in this direction was a recent paper by a subset of the current authors [FOCS 2013], where the result was established under the sub-optimal condition $\lambda < \lambda_c(\Delta + 1)$.

Our techniques also allow us to improve upon known bounds for decay of correlations for the hard core model on various regular lattices, including those obtained by Restrepo, Shin, Vigoda and Tetali [FOCS 11] for the special case of $\mathbb{Z}^2$ using sophisticated numerically intensive methods tailored to that special case.

1 Introduction

1.1 Background This paper studies the problem of approximately counting independent sets and matchings in sparse graphs. We consider these problems within the more general formalism of *spin systems*. In this setting, one first defines a natural probability distribution over *configurations* (e.g., independent sets or matchings) in terms of local *interactions*. The counting problem then corresponds to computing the normalization constant, known as the *partition function* in the statistical physics literature. The partition function can also be seen as a generating function of the combinatorial structures being considered and is an interesting graph polynomial in its own right.

The first model we consider is the so called *hard core model*, which is defined as follows. We start with a graph $G = (V, E)$, and specify a *vertex activity* or *fugacity* parameter $\lambda > 0$. The configurations of the hard core model are the independent sets of the graph, and the model assigns a *weight* $w(I) = \lambda^{|I|}$ to each independent set I in G. The weights in turn determine a natural probability distribution $\mu(I) = \frac{1}{Z} w(I)$ over the independent sets known as the *Gibbs distribution*. Here,

$$Z = Z(\lambda) := \sum_{I:\text{independent set}} w(I)$$

is the *partition function*. Clearly, the problem of counting independent sets is the special case $\lambda = 1$.

Our next model is the *monomer-dimer model*, which has as its configurations all matchings of a given graph $G = (V, E)$. For a specified *dimer activity* $\gamma > 0$, the model assigns a weight $w(M) = \gamma^{|M|}$ to each matching M of the graph. As before, the weights define the Gibbs distribution $\mu(M) = \frac{1}{Z} w(M)$ over

*Alistair Sinclair, Computer Science Division, UC Berkeley. Email: sinclair@cs.berkeley.edu. Supported in part by NSF grant CCF-1016896 and by the Simons Institute for the Theory of Computing.

†Piyush Srivastava, Center for the Mathematics of Information, Caltech. Email: piyushsriva@gmail.com. Supported by NSF grant CCF-1319745. This work was done while this author was a graduate student at UC Berkeley and was supported by NSF grant CCF-1016896.

‡Daniel Štefankovič, Department of Computer Science, University of Rochester. Email: stefanko@cs.rochester.edu. Supported in part by NSF grant CCF-1016896. Part of this work was done while this author was visiting the Simons Institute for the Theory of Computing.

§Yitong Yin, State Key Laboratory for Novel Software Technology, Nanjing University, China. Email: yinyt@nju.edu.cn. Supported by NSFC grants 61272081 and 61321491. Part of this work was done while this author was visiting UC Berkeley.

matchings, where

$$Z = Z(\gamma) := \sum_{M:\text{matching}} w(M)$$

is the partition function. The problem of counting matchings again corresponds to the special case $\gamma = 1$.

The problem of approximating the partition function has received much attention, both as a natural generalization of counting and because of its connections to sampling.[1] Recent progress in relating the complexity of approximating the partition function to phase transitions, which we now describe, has provided further impetus to this line of research.

The first such result was due to Weitz [31], who exploited the properties of the Gibbs measure of the hard core model on the infinite d-ary tree. It was well known that this model exhibits the following phase transition: there exists a *critical activity* $\lambda_c(d)$ such that the total variation distance between the marginal probability distributions induced at the root of the tree by *any* two fixings of the independent set on all the vertices at distance ℓ from the root decays exponentially in ℓ when $\lambda < \lambda_c(d) := \frac{d^d}{(d-1)^{d+1}}$, but remains bounded away from 0 even as $\ell \to \infty$ when $\lambda > \lambda_c(d)$. (The former condition is also referred to as *correlation decay*, since the correlation between the configuration at the root of the tree and a fixed configuration at distance ℓ from the root decays exponentially in ℓ; it is also called *spatial mixing*.) Weitz showed that for all $\lambda < \lambda_c(d)$ (i.e., in the regime where correlation decay holds on the d-ary tree), there exists a deterministic FPTAS for the partition function of the hard core model on all graphs of degree at most $d + 1$. (Note that the condition on λ is only in terms of the d-ary tree, while the FPTAS applies to all graphs.) This connection to phase transitions was further strengthened by Sly [26] (see also [7, 27]), who showed that an FPRAS for the partition function of the hard core model with $\lambda > \lambda_c(d)$ on graphs of degree $d + 1$ would imply NP = RP.

In addition to establishing a close connection between the complexity of a natural computational problem and an associated phase transition, Weitz's algorithm had the further interesting feature of not being based on Markov chain Monte Carlo (MCMC) methods; rather, it used a deterministic procedure based on proving that decay of correlations on the d-ary tree implies decay of correlations on all graphs of degree at most $d + 1$. To date, no MCMC algorithms are known for the approximation of the partition function of the hard core model on graphs of degree at most $d + 1$ which run in polynomial time for all $\lambda < \lambda_c(d)$.

Weitz's algorithm led to an exploration of his approach for other problems too. For example, in the case of the monomer-dimer model—unlike that of the hard core model—there does exists a *randomized* polynomial time algorithm (based on MCMC) for approximating the partition function which works for every $\gamma > 0$, without any bounds on the degree of the graph [12]. However, finding a deterministic algorithm for the problem remains open. Bayati, Gamarnik, Katz, Nair and Tetali [3] made progress on this question by showing that Weitz's approach could be used to derive a deterministic algorithm that runs in polynomial time for bounded degree graphs, and is sub-exponential on general graphs.

The algorithms of both Weitz and Bayati *et al.* are therefore polynomial time only on bounded degree graphs, and in particular, for a given value of the parameter λ (or γ in the case of the monomer-dimer model) the running time of these algorithms on graphs of maximum degree $d + 1$ depends upon the rate of decay of correlations on the infinite d-ary tree. Further, these results are obtained by showing that decay of correlations on the d-ary tree implies a similar decay on all graphs of maximum degree $d + 1$.

There are two important shortcomings of such results. First, in statistical physics one is often interested in special classes of graphs such as regular lattices. One can reasonably expect that the rate of decay of correlations on such graphs should be better than that predicted by their maximum degree. Second, these results have no non-trivial consequences even in very special classes of sparse unbounded degree graphs, such as graphs drawn from the Erdős-Rényi model $\mathcal{G}(n, d/n)$ for constant d.

This state of affairs leads to the following natural question: is there a finer notion of degree that can be used in these results in place of the maximum degree? Progress in this direction was made recently for the case of the hard core model in [25], where it was shown that one can get decay of correlation results in terms of the *connective constant,* a natural and well-studied notion of average degree. The connective constant of a regular lattice of degree $d+1$ is typically substantially less than d; and it is bounded even in the case of sparse random graphs such as those drawn from $\mathcal{G}(n, d/n)$, which have unbounded maximum degree. By analogy with the bounded degree case, one might hope to get correlation decay on graphs with connective constant at most Δ for all $\lambda < \lambda_c(\Delta)$. In [25], such a result was

[1] For the large class of self-reducible problems, it can be shown that approximating the partition function is polynomial-time equivalent to approximate sampling from the Gibbs distribution [13].

proven under the stronger condition $\lambda < \lambda_c(\Delta+1)$. The latter bound is tight asymptotically as $\Delta \to \infty$ (because $\lambda_c(\Delta+1)/\lambda_c(\Delta) \to 1$ as $\Delta \to \infty$), but is sub-optimal in the important case of small Δ.

1.2 Contributions In this paper, we show that one can indeed replace the maximum degree by the connective constant in the results of both Weitz [31] and Bayati *et al.* [3]. In particular, we show that for both the hard core and the monomer-dimer models, decay of correlations on the d-ary tree determines the rate of decay of correlations—as well as the complexity of deterministically approximating the partition function—in all graphs of *connective constant* at most d, without any dependence on the maximum degree. The specific notion of decay of correlations that we establish is known in the literature as *strong spatial mixing* [10, 18, 19, 31], and stipulates that the correlation between the state of a vertex v and another set S of vertices at distance ℓ from v should decay exponentially in ℓ even when one is allowed to fix the state of vertices close to v to arbitrary values (see the full version [23] for a precise definition). Prior to the role it played in the design of deterministic approximate counting algorithms in Weitz's work [31], strong spatial mixing was already a widely studied notion in computer science and mathematical physics for its utility in analyzing the mixing time of Markov chains [10, 18, 19], and hence an improved understanding of conditions under which it holds is of interest in its own right.

We now give an informal description of the connective constant [11, 17]; see Section 2.4 for precise definitions. Given a graph G and a vertex v in G, let $N(v,\ell)$ denote the number of self avoiding walks in G of length ℓ starting at v. A graph family $\mathcal{F}$ is said to have connective constant Δ if for all graphs in $\mathcal{F}$, the number of self-avoiding walks of length at most ℓ for large ℓ grows as Δ^ℓ, i.e., if $\ell^{-1}\log\sum_{i=1}^{\ell} N(v,i) \sim \log\Delta$ (the definition can be applied to both finite and infinite graphs; see Section 2.4). Note that in the special case graphs of maximum degree $d+1$, the connective constant is at most d. It can, however, be much lower that this crude bound: for any $\epsilon > 0$, the connective constant of graphs drawn from $G(n, d/n)$ is at most $d(1+\epsilon)$ with high probability (w.h.p.) (see, e.g., [23] for a proof), even though their maximum degree is $\Omega\left(\frac{\log n}{\log\log n}\right)$ w.h.p.

Our first main result can now be stated as follows.

Theorem 1.1 (Main, Hard core model). *Let $\mathcal{G}$ be a family of finite graphs of connective constant at most Δ, and let λ be such that $\lambda < \lambda_c(\Delta)$. Then there is an FPTAS for the partition function of the hard core model with vertex activity λ for all graphs in $\mathcal{G}$. Further, even if $\mathcal{G}$ contains locally finite infinite graphs, the model exhibits strong spatial mixing on all graphs in $\mathcal{G}$.*

Remark 1.1. In [25], the above result was proved under the stronger hypothesis $\lambda < \lambda_c(\Delta+1)$. The above result therefore subsumes the main results of [25]. It is also optimal in the following sense: there cannot be an FPRAS for graphs of connective constant at most Δ which works for $\lambda > \lambda_c(\Delta)$, unless NP = RP. This follows immediately from the hardness results for the partition function of the hard core model on bounded degree graphs [26, 27] since graphs of degree at most $d+1$ have connective constant at most d.

An immediate corollary of Theorem 1.1 is the following.

Corollary 1.2. *Let $\lambda < \lambda_c(d)$. Then, there is an algorithm for approximating the partition function of graphs drawn from $\mathcal{G}(n, d/n)$ up to a factor of $(1 \pm \epsilon)$ which, with high probability over the random choice of the graph, runs in time polynomial in n and $1/\epsilon$.*

A second consequence of Theorem 1.1 is a further improvement upon the spatial mixing bounds obtained in [25] for various lattices, as shown in Table 1. For each lattice, the table shows the best known upper bound for the connective constant and the strong spatial mixing (SSM) bounds we obtain using these values in Theorem 1.1. In the table, a value α in the "λ" column means that SSM is shown to hold for the appropriate lattice whenever $\lambda \le \alpha$. As expected, improvements over our previous results in [25] are the most pronounced for lattices with smaller maximum degree.

The table shows that except in the case of the 2D integer lattice $\mathbb{Z}^2$, our general result immediately gives improvements on the best known SSM bounds for all lattices using only previously known estimates of the connective constant. Not unexpectedly, our bound for $\mathbb{Z}^2$ using the connective constant as a black-box still improves upon Weitz's bound but falls short of the bounds obtained by Restrepo *et al.* [22] and Vera *et al.* [28] using numerically intensive methods tailored to this special case. However, as we noted in [25], any improvement in the bound on the connective constant would immediately yield an improvement in our SSM bound. Indeed, in the full version [23], we use a tighter analysis of the connective constant of a suitably constructed self-avoiding walk tree of $\mathbb{Z}^2$ to show that SSM holds on this lattice whenever $\lambda < 2.538$, which improves upon the specialized bound

$\lambda < 2.48$, obtained in the papers [22, 28]. We note that this improvement would not be possible using only our earlier results in [25].

Our second main result concerns the monomer-dimer model.

Theorem 1.3 (Main, Monomer-dimer model). *Let $\mathcal{G}$ be a family of finite graphs of connective constant at most Δ, and let $\gamma > 0$ be any fixed edge activity. Then there is an FPTAS for the partition function of the monomer-dimer model with edge activity γ for all graphs in $\mathcal{G}$. More specifically, the running time of the FPTAS for producing an $(1 \pm \epsilon)$ factor approximation is $(n/\epsilon)^{O(\sqrt{\gamma\Delta}\log\Delta)}$.*

The previous best deterministic approximation algorithm for the partition function of the monomer-dimer model was due to Bayati *et al.* [3], and ran in time $(n/\epsilon)^{O(\sqrt{\gamma d}\log d)}$ for graphs of degree at most $d+1$. Thus, our algorithm replaces the maximum degree constraint of Bayati *et al.* by a corresponding constraint on the connective constant, without requiring any bounds on the maximum degree. In particular, for graphs such as $\mathcal{G}(n, d/n)$ which have bounded connective constant and unbounded degree, our analysis yields a polynomial time algorithm (for any fixed value of the edge activity γ) in contrast to the sub-exponential time algorithm obtained by Bayati *et al.* [3]. Using an observation of Kahn and Kim [14], Bayati *et al.* also pointed out that the $\sqrt{d}$ factor in the exponent of their running time was optimal for algorithms which are based on Weitz's framework and which use only the fact that the maximum degree of the graph is at most $d+1$. A similar observation shows that the $\sqrt{\gamma\Delta}$ factor in the exponent of our running time is optimal for algorithms in the Weitz framework which use bounds on the connective constant (see the the full version [23] for a more detailed discussion of this point). As an aside, we also note that when no bounds on the connective constant are available our FPTAS degrades to a sub-exponential algorithm, as does the algorithm of Bayati *et al.* in the case of unbounded degree graphs.

1.3 Techniques The analyses by Weitz [31] and Bayati *et al.* [3] both begin with the standard observation that obtaining an FPTAS for the marginal probabilities of the Gibbs distribution is sufficient in order to obtain an FPTAS for the partition function. The next non-trivial step is to show that this computation of marginal probabilities at a given vertex v in a graph G can be carried out on the tree $T_{SAW}(v, G)$ of *self-avoiding walks* in G starting at v. Transferring the problem to a tree allows one to write down a recurrence for the marginal probabilities of a node in the tree in terms of the marginal probabilities of its children. However, since the tree is of exponential size, one needs to truncate the tree at a small (logarithmic) depth in order to obtain a polynomial time algorithm. Such a truncation in turn introduces an "error" at the leaves. The challenge then is to show that this error contracts exponentially as the recurrence works its way up to the root.

The approach of [3, 31] (and similar results in [15, 16, 24]) for establishing this last condition takes the following general form: one shows that at each step of the recurrence, the correlation decay condition implies that the error at the parent node is less than a constant factor (less than 1) times the maximum (ℓ_∞ norm) of the errors at the children of the node. Intuitively, this strategy loses information about the structure of the tree by explaining the error at the parent in terms of *only one* of its children, and hence it is not surprising that the results obtained from it are only in terms of a local parameter such as the maximum degree.

The main technical contribution of [25] was to show that, by analyzing the decay in terms of the ℓ_2 norm—rather than the ℓ_∞ norm—of the errors at the children, one can get past this limitation and obtain a result in terms of the connective constant. Nevertheless, as stated above, the results obtained in [25] did not hold over the best possible range of parameters. Our main innovation in the present paper is to analyze instead a norm *adapted to* the parameters of the model, rather than a fixed norm such as ℓ_2 or ℓ_∞. Specifically, we show that optimal results can be obtained by analyzing the decay in terms of a carefully picked ℓ_q norm where q is chosen as a function of the connective constant and the model parameters (the fugacity λ in the hard core model and the edge activity γ in the monomer-dimer model). At a technical level, the use of these adaptive norms implies that we can no longer employ the relatively simpler convexity arguments used in [25] in order to bound the propagation of errors; characterizing the "worst case" error vectors now requires solving a more involved optimization problem, which is the main new technical challenge in this paper. In Section 3, we give a general framework for tackling this problem. Our model specific main results are then obtained as direct applications of this framework. We conjecture that our framework may find applications to other approximate counting problems as well.

1.4 Related work Approximating the partition function has traditionally been studied in the framework of Markov Chain Monte Carlo (MCMC) methods. For the hard core model on bounded degree graphs,

Lattice	Max. degree	Previous SSM bound λ	Connective Constant Δ	SSM bound in [25] λ	Our SSM bound λ
$\mathbb{T}$	6	0.762 [31]	4.251 419 [1]	0.937	0.961
$\mathbb{H}$	3	4.0 [31]	1.847 760 [4]	4.706	4.976
$\mathbb{Z}^2$	4	2.48 [22,28]	2.679 193 [21]	2.007	2.082 (2.538*)
$\mathbb{Z}^3$	6	0.762 [31]	4.7387 [21]	0.816	0.822
$\mathbb{Z}^4$	8	0.490 [31]	6.8040 [21]	0.506	0.508
$\mathbb{Z}^5$	10	0.360 [31]	8.8602 [21]	0.367	0.367
$\mathbb{Z}^6$	12	0.285 [31]	10.8886 [30]	0.288	0.288

* See the full version [23] for a description of how this improved bound is obtained.

Table 1: Strong spatial mixing bounds for various lattices. ($\mathbb{Z}^D$ is the D-dimensional Cartesian lattice; $\mathbb{T}$ and $\mathbb{H}$ denote the triangular and honeycomb lattices respectively.)

this line of work culminated in papers by Dyer and Greenhill [5] and Vigoda [29], who gave MCMC based FPRASs for $\lambda < 2/(d-1)$ for graphs of maximum degree at most $d+1$. Weitz [31] (see also [2]) introduced a new paradigm by using correlation decay directly to design a *deterministic* FPTAS and gave an algorithm under the condition $\lambda < \lambda_c(d)$ for graphs of degree at most $d+1$; this range of applicability was later proved to be optimal by Sly [26] (see also [7,27]). To date, no MCMC based algorithm is known to have a range of applicability as wide as Weitz's algorithm. Restrepo, Shin, Tetali, Vigoda and Yang [22] extended the approach using sophisticated computational methods to prove that spatial mixing for the hard core model holds on $\mathbb{Z}^2$ whenever $\lambda < 2.38$. However, none of the above mentioned results could handle even special classes of unbounded degree graphs. Thus, the problem of sampling from the hard core model on graphs drawn from $G(n, d/n)$ (which have unbounded degree) was studied using MCMC methods by Mossel and Sly [20] and more recently by Efthymiou [6], who gave a fast MCMC based sampler for $\lambda < 1/(2d)$. For a more detailed discussion of earlier work on the hard core mode we refer to the full version [23]. We note in passing that Weitz's approach has since been used to attack other approximate counting problems as well (see, e.g., [3,8]).

Sinclair *et al.* [25] considered the hard-core model on graphs of bounded connective constant Δ, and gave an FPTAS under the condition $\lambda < \lambda_c(\Delta+1)$. Since $\lambda_c(\Delta+1) > e/\Delta$, this also gave a polynomial time algorithm for sampling from the hard core model on $G(n, d/n)$ when $\lambda < e/d$. They also applied their techniques to study spatial mixing on regular lattices, and obtained improvements on previous results in all cases except for the case of $\mathbb{Z}^2$, where their results fell short of the specialized analysis of Restrepo *et al.* [22].

In contrast to the case of the hard core model, where algorithms with the largest range of applicability are already deterministic, much less is known about deterministic approximation of the partition function of the monomer-dimer model. Jerrum and Sinclair [12] gave an MCMC-based randomized algorithm which runs in polynomial time on all graphs (without any bounds on the maximum degree) for any fixed value of the edge activity λ. However, no deterministic algorithms with this range of applicability are known, even in the case of specific graph families such as $\mathcal{G}(n, d/n)$. So far, the best result in this direction is due to Bayati, Gamarnik, Katz, Nair and Tetali [3], whose algorithm produces a $(1 \pm \epsilon)$ factor approximation for the monomer-dimer partition function in time $(n/\epsilon)^{\tilde{O}(\sqrt{\gamma d})}$ on graphs of maximum degree d. Their result therefore yields a super-polynomial (though sub-exponential) algorithm in the case of graphs such as those drawn from $\mathcal{G}(n, d/n)$ (note that the results of this paper imply a polynomial time algorithm in this setting).

2 Preliminaries

Note: In the conference version, we focus on the monomer-dimer model. Technical details of the results on the hard core model can be found in the full version [23].

2.1 The monomer-dimer model As stated in the introduction, algorithms based on the Weitz framework derive an FPTAS for the partition function by giving an FPTAS for appropriate marginal probabilities; the latter reduction is based on standard "self-reducibility" arguments, which we defer to Appendix A. We now set up notation for the marginals.

Definition 2.1 (**Monomer probability**). Consider the Gibbs distribution of the monomer-dimer model

with dimer activity γ on a finite graph $G = (V, E)$, and let v be a vertex in V. We define the *monomer probability* $p(v, G)$ as $p_v := \mathbb{P}[v \notin M]$ which is the probability that v is unmatched (i.e., a *monomer*) in a matching M sampled from the Gibbs distribution.

Remark 2.1. The monomer-dimer model is often described in the literature in terms of a *monomer activity* λ instead of the *dimer activity* γ used here. In this formulation, the weight of a matching M is $\lambda^{u(M)}$, where $u(M)$ is the number of unmatched vertices (*monomers*) in M. The two formulations are equivalent: with dimer activity γ corresponding to monomer activity $\lambda = \frac{1}{\gamma^2}$.

2.2 Truncated recurrences with initial conditions As in the case of other correlation decay based algorithms (e.g., in [8, 31]), we will need to analyze recurrences for marginals on rooted trees with various initial conditions. We therefore set up some notation for describing such recurrences. For a vertex v in a tree T, we will denote by $|v|$ the distance of v from the root of the tree. Similarly, for a set S of vertices, $\delta_S := \min_{v \in S} |v|$.

Definition 2.2 (Cutset). Let T be any tree rooted at ρ. A *cutset* C is a set of vertices in T satisfying the following two conditions: (1) any path from ρ to a leaf v with $|v| \geq \delta_C$ must pass through C, and (2) C is an antichain, i.e., for any vertices u and v in C, neither vertex is an ancestor of the other in T. For a cutset C, we will denote by $T_{\leq C}$ the subtree of T obtained by removing the descendants of vertices in C from T.

A trivial example of a cutset is the set L of all the leaves of T. Another example we will often need is the set S_ℓ of all vertices at distance ℓ from ρ in T.

Definition 2.3 (Initial condition). An *initial condition* $\sigma = (S, P)$ is a set S of vertices in T along with an assignment $P : S \to [0, b]$ of bounded positive values to vertices in S.

We are now ready to describe the tree recurrences. Given an initial condition $\sigma = (S, P)$ along with a default value b_0 for the leaves, a family of functions $f_d : [0, b]^d \to [0, b]$ for every positive integer $d \geq 1$, and a vertex u in T, we let $F_u(\sigma)$ denote the value obtained at u by iterating the tree recurrences f on the subtree T_u rooted at u under the initial condition σ. Formally, we define $F_u(\sigma) = b_0$ when $u \notin S$ is a leaf, and

$$F_u(\sigma) = \begin{cases} P(u) & \text{when } u \in S, \\ f_d\left(F_{u_1}(\sigma), \ldots, F_{u_d}(\sigma)\right) & \text{when } u \notin S \text{ is of arity } d \geq 1 \text{ and has children } u_1, u_2, \ldots, u_d. \end{cases} \tag{2.1}$$

2.3 The self-avoiding walk tree and associated recurrences Given a vertex v in a graph G, one can define a rooted tree $T_{SAW}(v, G)$ of self-avoiding walks (called the *self-avoiding walk tree*, or *SAW tree*) starting at v, as follows: the root of the tree represents the trivial self-avoiding walk that ends at v, and given any node u in the tree, its children represent all possible self-avoiding walks than can be obtained by extending the self-avoiding walk represented by u by exactly one step. The importance of the self-avoiding walk tree for computation stems from the beautiful results of Godsil [9] (for the monomer-dimer model) and Weitz [31] (for the hard core model), which allow the derivation of simple recurrences for the monomer probability $p_v(G)$ on general graphs. We defer the discussion of Weitz's reduction to the full version [23], and concentrate here on the monomer-dimer model.

Theorem 2.1 (Godsil [9]). *Let v be a vertex in a graph G, and consider the monomer-dimer model with dimer activity $\gamma > 0$ on the graphs G and $T_{SAW}(v, G)$. We then have $p_v(G) = p_v(T_{SAW}(v, G))$.*

The promised recurrence for $p_v(G)$ can now be derived using dynamic programming on the tree $T_{SAW}(v, G)$. In particular, let T be any tree rooted at ρ, and let ρ_i, $1 \leq i \leq d$ be the children of ρ. Denoting by p_i the monomer probability $p_{\rho_i}(T_{\rho_i})$ at the root of the subtree T_{ρ_i}, one can then show that (see, e.g., [14])

$$p_\rho(T) = f_{d,\gamma}(p_1, p_2, \ldots, p_d) := \frac{1}{1 + \gamma \sum_{i=1}^{d} p_i}. \tag{2.2}$$

In terms of our notation for tree recurrences, we note that the actual computation of $p_\rho(T)$ corresponds to computing $F_\rho(\mathbf{1}_L)$, where the initial condition $\mathbf{1}_L$ assigns the value 1 to all vertices in L, the cutset comprising all the leaves (and with the boundary value b_0 set to 1), since the base case of the recurrence comprises a single vertex which has monomer probability 1 by definition.

Note that the self-avoiding tree can be of exponential size, so that Godsil's reduction does not immediately yield an efficient algorithm for computing $p_\rho(G)$. In order to obtain an algorithm, we would need to consider truncated versions of the recurrence,

obtained by specifying initial conditions on the cut-set S_ℓ comprising all vertices at distance ℓ from ρ. Since $f_{d,\gamma}$ is monotonically decreasing in each of its arguments, we have

(2.3)
$$F_\rho(\mathbf{0}_\ell) \le p_\rho(T) \le F_\rho(\mathbf{1}_\ell) \quad \text{when } \ell \text{ is even, and}$$
$$F_\rho(\mathbf{0}_\ell) \ge p_\rho(T) \ge F_\rho(\mathbf{1}_\ell) \quad \text{when } \ell \text{ is odd.}$$

Here, the initial condition $\mathbf{0}_\ell$ (respectively, $\mathbf{1}_\ell$) assigns the value 0 (respectively, 1) to every vertex in S_ℓ. Given these conditions, it is sufficient to show that the difference between $F_\rho(\mathbf{0}_\ell)$ and $F_\rho(\mathbf{1}_\ell)$ decreases exponentially in ℓ in order to establish that truncated versions of the recurrence converge to the true answer $p_\rho(T)$ exponentially fast in the "truncation length" ℓ.

2.4 The connective constant We now recall the definition of the connective constant of a graph. Given a vertex v in a locally finite graph, let $N(v,l)$ be the number of self-avoiding walks of length l in the graph which start at v. The *connective constant* $\Delta(G)$ of an infinite graph G is then defined as $\sup_{v\in V} \limsup_{\ell\to\infty} N(v,\ell)^{1/\ell}$. Note that for the special class of vertex-transitive graphs (such as Cartesian lattices), the supremum over v in the definition is clearly not required, and further, the lim sup can be replaced by a limit [17].

The definition was extended in [25] to families of finite graphs parametrized by size. As observed there, such a parametrization is natural for algorithmic applications.

Definition 2.4 (Connective constant: finite graphs [25]). Let $\mathcal{F}$ be a family of finite graphs. The connective constant of $\mathcal{F}$ is at most Δ if there exist constants a and c such that for any graph $G = (V,E)$ in $\mathcal{F}$ and any vertex v in G, we have $\sum_{i=1}^{\ell} N(v,i) \le c\Delta^\ell$ for all $\ell \ge a \log |V|$.

As observed earlier, the connective constant of a graph of maximum degree $d+1$ is at most d, but can be much smaller than this crude bound. For example, though the maximum degree of a graph drawn from the Erdős–Rényi model $\mathcal{G}(n, d/n)$ is $\Theta(\log n/\log\log n)$ w.h.p, it is not hard to show (see, e.g., [23]) that for any fixed $\epsilon > 0$, the connective constant of such a graph is at most $d(1+\epsilon)$ w.h.p.

Remark 2.2. Note that the connective constant has a natural interpretation as the "average arity" of the SAW tree, since vertices in $T_{SAW}(v,G)$ at distance ℓ from the root are in bijection with self-avoiding walks of length ℓ starting at v.

3 Decay of correlations on the SAW tree

In this section, we lay the groundwork for proving decay of correlations results for the tree recurrences F_ρ defined in eq. (2.1) for both the hard core and monomer-dimer models: such a result basically affirms that truncating the recurrence at a small depth ℓ is sufficient in order to approximate F_ρ with good accuracy. Our proof will use the *message approach* [16, 22, 24], which proceeds by defining an appropriate function ϕ of the marginals being computed and then showing a decay of correlation result for this function.

Definition 3.1 (Message [16, 22, 24]). Given a positive real number b, a *message* is a strictly increasing and continuously differentiable function $\phi : (0,b] \to \mathbb{R}$, with the property that the derivative of ϕ is bounded away from 0 on its domain. A message ϕ is guaranteed to admit a continuously differentiable inverse, which we will denote by ψ.

In the rest of this section, we will work in the abstract framework described in Section 2.2, to illustrate how the message approach can be used to get strengthened decay of correlation estimates as compared to those obtained from direct analyses of one step of the recurrence. We will then instantiate our framework with an appropriately chosen message for the monomer-dimer model in Section 4. Its application to the hard core model differs only in the use of a different message and can be found in the full version [23].

We begin by fixing the boundary value b_0 for the leaves in our recurrence framework, and assume that the initial conditions specify values in the interval $[0,b]$. We assume that we have a set of tree recurrences $f_d : [0,b]^d \to [0,b]$ for every positive integer $d \ge 1$. The only constraints we put on the recurrences in this section are the following (both of which are trivially satisfied by the recurrences for the hard core and the monomer-dimer model).

Condition 3.1 (Consistency). *We say that a set of recurrences $\{f_d\}_{d\ge1}$, where f_d is d-variate, are* consistent *if they obey the following two conditions: (1) if $\boldsymbol{x} \in \mathbb{R}^d$ is a permutation of $\boldsymbol{y} \in \mathbb{R}^d$, then $f_d(\boldsymbol{x}) = f_d(\boldsymbol{y})$; and (2) if all but the first k co-ordinates of $\boldsymbol{x} \in \mathbb{R}^d$ are 0, then $f_d(\boldsymbol{x}) = f_k(x_1, x_2, x_3, \ldots, x_k)$.*

Given the message ϕ (and its inverse ψ), we further define f_d^ϕ by

$$f_d^\phi(x_1, x_2, \ldots, x_d) := \phi\left(f_d\left(\psi(x_1), \psi(x_2), \ldots, \psi(x_d)\right)\right).$$

We then have the following simple consequence of the mean value theorem (a proof can be found in Appendix B).

Lemma 3.2 (**Mean value theorem**). *Consider two vectors* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *in* $\phi([0,B])^d$. *Then there exists a vector* $\boldsymbol{z} \in [0,\infty)^d$ *such that*

(3.4)
$$\left|f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y})\right| \le \Phi\left(f_d(\boldsymbol{z})\right) \sum_{i=1}^d \frac{|y_i - x_i|}{\Phi(z_i)} \left|\frac{\partial f_d}{\partial z_i}\right|,$$

where $\Phi := \phi'$ *is the derivative of* ϕ, *and by a slight abuse of notation we denote by* $\frac{\partial f_d}{\partial z_i}$ *the partial derivative of* $f_d(R_1, R_2, \ldots, R_d)$ *with respect to* R_i *evaluated at* $\boldsymbol{R} = \boldsymbol{z}$.

The first step of our approach is similar to that taken in the papers [15, 16, 22, 24] in that we will use an appropriate message—along with the estimate in Lemma 3.2—to argue that the "distance" between two input message vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ at the children of a vertex shrinks by a constant factor at each step of the recurrence. Previous works [15, 16, 22, 24] showed such a decay on some version of the ℓ_∞ norm of the "error" vector $\boldsymbol{x} - \boldsymbol{y}$: this was achieved by bounding the appropriate dual ℓ_1 norm of the gradient of the recurrence. Our intuition is that in order to achieve a bound in terms of a global quantity such as the connective constant, it should be advantageous to use a more global measure of the error such as an ℓ_q norm for some $q < \infty$.

In line with the above plan, we will attempt to bound the right hand side of eq. (3.4) in terms of $\|\boldsymbol{x} - \boldsymbol{y}\|_q$ for an appropriate value of $q < \infty$ by maximizing the sum while keeping $f_d(\boldsymbol{z})$ fixed. A similar approach was taken by Sinclair *et al.* [25], who carried out the above maximization using relatively simple concavity arguments. However, their arguments worked only for $q = 2$, and it was this restriction that led to their results being sub-optimal. The main technical innovation of this paper is to get past the requirement $q = 2$ using a more flexible optimization than that used in [25]. To do this, we will seek to establish the following property for our messages (the exponent a will be the Hölder conjugate of the value of q that we eventually use).

Definition 3.2. Given a consistent family of recurrences $\{f_d\}_{d\ge1}$, a message ϕ (with $\Phi := \phi'$) is said to be *symmetrizable with exponent* a with respect to the family if it satisfies the following two conditions:

1. Let $\mathcal{D}$ be the domain of the recurrence family. For every positive integer d and every real $B > 0$ for which the program
$$\begin{aligned} \max \quad & \sum_{i=1}^d \left(\frac{1}{\Phi(x_i)}\left|\frac{\partial f_d}{\partial x_i}\right|\right)^a, \qquad \text{where} \\ & f_d(\boldsymbol{x}) = B \\ & x_i \in \mathcal{D}, \qquad 1 \le i \le d \end{aligned}$$
is feasible, it also has a solution $\boldsymbol{x}$ in which all the non-zero entries of $\boldsymbol{x}$ are equal. (We assume implicitly that $0 \in \mathcal{D}$.)
2. $\lim_{x_i \to 0^+} \frac{1}{\Phi(x_i)}\left|\frac{\partial f_d}{\partial x_i}\right| = 0$ for all $d \ge 1$, and for any fixed values of the x_j, $j \ne i$.

For symmetrizable messages, we will be able to bound the quantity $|f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y})|$ in terms of $\|\boldsymbol{x} - \boldsymbol{y}\|_q$, where $1/a + 1/q = 1$, and our improved correlation decay bounds will be based on the fact that symmetrizability can be shown to hold under a wider range of values of q than that required by the concavity conditions used in [25]. Our bounds will be stated in terms of the following notion of decay.

Notation. Given a d-variate function f_d and a scalar x, we denote by $f_d(x)$ the quantity $f_d(x, x, \ldots, x)$.

Definition 3.3 (**Decay factor** α). Let ϕ be a message with derivative Φ, and let a and q be positive reals such that $\frac{1}{a} + \frac{1}{q} = 1$. We define the functions $\Xi_{\phi,q}(d,x)$ and $\xi_{\phi,q}(d)$ as follows:
$$\Xi_{\phi,q}(d,x) := \frac{1}{d}\left(\frac{\Phi(f_d(x))\,|f_d'(x)|}{\Phi(x)}\right)^q;$$
$$\xi_{\phi,q}(d) := \sup_{x\ge0} \Xi_{\phi,q}(d,x).$$

The *decay factor* α is then defined as

(3.5)
$$\alpha := \sup_{d\ge1} \xi_{\phi,q}(d).$$

Armed with the above definitions, we are now ready to prove Lemma 3.3, which provides the requisite decay bound for one step of the tree recurrence. The main technical step in applying this lemma is to find a, q as in the definition and a message ϕ symmetrizable with exponent a for which the decay factor α is small; Lemma 3.4 below then shows how the decay factor comes into play in proving exponential decay of correlations over the tree.

Lemma 3.3. *Let* ϕ *be a message with derivative* Φ, *and let* a *and* q *be positive reals such that* $\frac{1}{a} + \frac{1}{q} = 1$. *If* ϕ *is symmetrizable with exponent* a, *then for any two vectors* $\boldsymbol{x}, \boldsymbol{y}$ *in* $\phi([0,b])^d$, *there exists an integer* $k \le d$ *such that*
$$\left|f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y})\right|^q \le \xi_{\phi,q}(k)\|\boldsymbol{x} - \boldsymbol{y}\|_q^q.$$

Proof. We apply Lemma 3.2. Assuming $\boldsymbol{z}$ is as defined in that lemma, we have by Hölder's inequality

$$\begin{aligned}\left|f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y})\right| &\le \Phi(f_d(\boldsymbol{z}))\sum_{i=1}^{d}\frac{|y_i - x_i|}{\Phi(z_i)}\left|\frac{\partial f_d}{\partial z_i}\right| \\ &\le \Phi(f_d(\boldsymbol{z}))\|\boldsymbol{x}-\boldsymbol{y}\|_q \\ &\quad\cdot\left(\sum_{i=1}^{d}\left(\frac{1}{\Phi(z_i)}\left|\frac{\partial f_d}{\partial z_i}\right|\right)^a\right)^{1/a}.\end{aligned}$$

Since ϕ is symmetrizable with exponent a, we can replace $\boldsymbol{z}$ in the above inequality with a vector $\tilde{\boldsymbol{z}}$ all of whose non-zero entries are equal to some fixed real $\tilde{z}$. Let $k \le d$ be the number of non-zero entries in $\tilde{\boldsymbol{z}}$. Using the consistency condition, we then get

$$\begin{aligned}\left|f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y})\right| &\le \Phi(f_k(\tilde{z}))\|\boldsymbol{x}-\boldsymbol{y}\|_q \\ &\quad\cdot\left(\sum_{i=1}^{k}\left(\frac{1}{k\Phi(\tilde{z})}\left|f_k'(\tilde{z})\right|\right)^a\right)^{1/a} \\ &= \frac{1}{k^{1-1/a}}\frac{\Phi(f_k(\tilde{z}))\left|f_k'(\tilde{z})\right|}{\Phi(\tilde{z})}\|\boldsymbol{x}-\boldsymbol{y}\|_q.\end{aligned}$$

Raising both sides to the power q, and using $\frac{1}{a}+\frac{1}{q}=1$ and the definitions of the functions Ξ and ξ, we get the claimed inequality. □

Given a message ϕ satisfying the conditions of Lemma 3.3, we can easily prove the following lemma on the propagation of errors in locally finite infinite trees. Recall that $F_\rho(\sigma)$ denotes the value computed by the recurrence at the root ρ under an initial condition σ. The lemma quantifies the dependence of $F_\rho(\sigma)$ on initial conditions σ which are fixed everywhere except at some cutset C, in terms of the distance of C from ρ.

Lemma 3.4. *Let T be a finite tree rooted at ρ. Let C be a cutset in T at distance at least 1 from the root which does not contain any leaves, and let C' be the cutset consisting of the children of vertices in C. Consider two arbitrary initial conditions σ and τ on $T_{\le C'}$ which differ only on C', and which assign values from the interval $[0,b]$. Given a recurrence family $\{f_d\}_{d\ge1}$, let a and q be positive reals such that $\frac{1}{a}+\frac{1}{q}=1$ and suppose ϕ is a message that is symmetrizable with exponent a. We then have*

$$|F_\rho(\sigma) - F_\rho(\tau)|^q \le \left(\frac{M}{L}\right)^q \sum_{v\in C}\alpha^{|v|},$$

where α is as defined in eq. (3.5), and L and M are defined as follows:

$$L := \inf_{x\in(0,b)}\phi'(x);\quad M := \max_{v\in C}\left|\phi(F_v(\sigma)) - \phi(F_v(\tau))\right|.$$

For a proof of this lemma, see Appendix B.

4 A message for the monomer-dimer model

In this section, we apply the general framework of Section 3 to the monomer-dimer model. As in the case of the hard core model, the first step is to choose an appropriate message. Unfortunately, unlike the case of the hard core model where we can show that an already known message is sufficient (see the full version [23] for details), we need to find a new message function in this case. We claim that the following message works:

$$\phi(x) := \frac{1}{2}\log\left(\frac{x}{2-x}\right), \text{ so that} \tag{4.6}$$
$$\Phi(x) := \phi'(x) = \frac{1}{x(2-x)}.$$

Note that ϕ is strictly increasing and continuously differentiable on the interval $(0,1]$, and its derivative is bounded away from 0 on that interval. Thus, ϕ satisfies the conditions required in the definition of a message (note that the bound b used in the definition is 1 in the case of the monomer-dimer model). Now, in order to apply Lemma 3.4, we first study the symmetrizability of ϕ in the following technical lemma.

Lemma 4.1. *Fix $r \in (1,2]$. The message ϕ as defined in eq. (4.6) is symmetrizable with exponent r with respect to the tree recurrences $\{f_{d,\gamma}\}_{d\ge1}$ of the monomer-dimer model.*

We defer the proof of the above lemma to Appendix C. As discussed above, we will need to choose the exponent r carefully in terms of the parameters in order to obtain an optimal decay factor. We begin with a technical lemma which characterizes the behavior of the function ξ used in the definition of the decay factor. For ease of notation, we drop the subscript ϕ from the notation for ξ.

Lemma 4.2. *Consider the monomer-dimer model with edge activity γ, and let ϕ be the message chosen in (4.6). For any $q>1$, we have $\xi_q(d) = \Xi_q(d, \tilde{p}_\gamma(d))$, where $\tilde{p}_\gamma(d)$ satisfies $\Xi_q^{(0,1)}(d,\tilde{p}_\gamma(d)) = 0$ and is given by*

$$\tilde{p}_\gamma(d) := \frac{\sqrt{1+4\gamma d}-1}{2\gamma d}.$$

Proof. Plugging in Φ from eq. (4.6) in the definition of Ξ, we get

$$\begin{aligned}\Xi_q(d,x) &= d^{q-1}\left(\frac{\gamma x(2-x)f_{d,\gamma}(x)}{2-f_{d,\gamma}(x)}\right)^q \\ &= d^{q-1}\left(\frac{\gamma x(2-x)}{1+2\gamma dx}\right)^q, \\ &\qquad \text{since } f_{d,\gamma}(x) = \frac{1}{1+\gamma dx}.\end{aligned}$$

Taking the partial derivative with respect to the second argument, we get

$$\Xi_q^{(0,1)}(d,x) = \frac{2q\Xi_q(d,x)}{x(2-x)(1+2\gamma dx)}\left[1-x-\gamma dx^2\right].$$

For fixed d, and $0 \le x \le 1$, the quantity outside the square brackets is always positive, while the expression inside the square brackets is strictly decreasing in x. Thus, any zero of the expression in the brackets in the interval $[0,1]$ will be a unique maximum of Ξ_q. By solving the quadratic, we see that $\tilde{p}_\gamma(d)$ as defined above is such a solution. Thus, $\Xi_q(d,x)$ is maximized at $\tilde{p}_\gamma(d)$ as defined above, and hence $\xi_q(d) = \Xi_q(d, \tilde{p}_\gamma(d))$. □

Given the edge activity γ and an upper bound Δ on the connective constant of the graph family being considered, we now choose $D > \max(\Delta, 3/(4\gamma))$. We claim that we can get the required decay factor by choosing

(4.7)
$$\frac{1}{r} = 1 - \frac{1}{\sqrt{1+4\gamma D}}; \qquad \frac{1}{q} = 1 - \frac{1}{r} = \frac{1}{\sqrt{1+4\gamma D}}.$$

Note that the choice of D implies that $1 < r \le 2$, so that ϕ is symmetrizable with respect to r. The following lemma shows that this choice of r indeed gives us the required decay factor. We emphasize the dependence of the decay factor on the model parameters by setting $\nu_\gamma(d) := \xi_q(d)$, where q is as chosen in eq. (4.7).

Lemma 4.3. *Fix $\gamma > 0$ and $D > 3/4\gamma$, and let q be as chosen in (4.7). Then the function $\nu_\gamma : \mathbb{R}^+ \to \mathbb{R}^+$ is maximized at $d = D$. Further, the decay factor α is given by*

$$\alpha = \nu_\gamma(D) = \frac{1}{D}\left(1 - \frac{2}{1+\sqrt{1+4\gamma D}}\right)^q.$$

Proof. We consider the derivative of $\nu_\gamma(d)$ with respect to d. Recalling that $\nu_\gamma(d) = \xi(d) = \Xi(d, \tilde{p}_\gamma(d))$ and using the chain rule, we have

$$\begin{aligned}
\nu'_\gamma(d) &= \Xi^{(1,0)}(d,\tilde{p}) + \Xi^{(0,1)}(d,\tilde{p})\frac{\mathrm{d}\tilde{p}}{\mathrm{d}d} \\
&= \Xi^{(1,0)}(d,\tilde{p}), \\
&\qquad \text{since } \Xi^{(0,1)}(d,\tilde{p}) = 0 \text{ by definition of } \tilde{p} \\
&= \frac{\Xi(d,\tilde{p})}{d(1+2\gamma d\tilde{p})}\left[q - 1 - 2\gamma d\tilde{p}\right] \\
&= \frac{\Xi(d,\tilde{p})}{d(1+2\gamma d\tilde{p})}\left[\sqrt{1+4\gamma D} - \sqrt{1+4\gamma d}\right], \qquad (4.8)
\end{aligned}$$

where we in the last line we substitute the values $\tilde{p}_\gamma(d) = (\sqrt{1+4\gamma d}-1)/(2\gamma d)$ from Lemma 4.2 and $q = \sqrt{1+4\gamma D}$ from eq. (4.7). Now, we note that in eq. (4.8), the quantity outside the square brackets is always positive, while the quantity inside the square brackets is a strictly decreasing function of d which is positive for $d < D$ and negative for $d > D$. It follows that $\nu'_\gamma(d)$ has a unique zero at $d = D$ for $d \ge 0$, and this zero is a global maximum of ν_γ. □

We are now ready to prove our main result for the monomer-dimer model, Theorem 1.3 from the introduction. Given Lemmas 3.4 and 4.3, only some standard arguments are needed to finish the proof.

Proof of Theorem 1.3. Let $\mathcal{F}$ be any family of finite graphs with connective constant at most Δ. Given the vertex activity γ of the monomer-dimer model, we choose $D = \max(\Delta, 3/(4\gamma))$. Using Lemma 4.3, we then see that the decay factor α appearing in Lemma 3.4 can be chosen to be $\alpha = [1 - 2/(1+\sqrt{1+4\gamma D})]^q/D$. Now, let G be any graph (with n vertices) from $\mathcal{F}$, and let v be a vertex in G. As observed in Section 2.1, it is sufficient to construct an FPTAS for $p_v(G)$ in order to derive an FPTAS for the partition function.

Consider the self-avoiding walk tree $T_{SAW}(v,G)$ rooted at v (as defined in Section 2.1). From Godsil's theorem (Theorem 2.1), we know that $p_v(G) = p_v(T_{SAW}(v,G))$. Let C_ℓ denote the cutset in $T_{SAW}(v,G)$ consisting of all vertices at distance ℓ from v. Since $\mathcal{F}$ has connective constant at most Δ, there exist constants a and c such that if $\ell \ge a\log n$, we have $\sum_{i=1}^{\ell} N(v,\ell) \le c\Delta^\ell$. We will now apply Lemma 3.4 with q as defined in eq. (4.7). We first observe that the quantities L and M in the lemma can be taken to be

$$L = 1, \quad \text{and,} \quad M = \frac{1}{2}\log(1+2\gamma n),$$

since the degree of any vertex in G is at most n.[3] Define $c_0 := (M/L)^q$. Starting with the result of Lemma 3.4, we then use $|C_\ell| \le c\Delta^\ell$, substitute the value of α, and finally employ $D \ge \Delta$ to get

$$\begin{aligned}
|F_v(\mathbf{0}_\ell) - F_v(\mathbf{1}_\ell)|^q &\le c_0 \sum_{u\in C_\ell} \alpha^\ell \le c\cdot c_0 \cdot (\alpha\Delta)^\ell, \\
&\le c\cdot c_0 \cdot \left(1 - \frac{2}{1+\sqrt{1+4\gamma D}}\right)^{q\ell}, \qquad (4.9)
\end{aligned}$$

[3] Since the degree of every vertex v in the graph is n, every boundary condition sigma satisfies $1 \ge F_v(\sigma) \ge \frac{1}{1+\gamma n}$. Substituting these bounds in the definition of M in Lemma 3.4 yields the claimed bound.

Raising both sides to the power $1/q$ and substituting for c_0 and q, we then have

(4.10)
$$|F_v(\mathbf{0}_\ell) - F_v(\mathbf{1}_\ell)| \le \frac{1}{2} c^{1/\sqrt{1+4\gamma D}} \cdot \log(1+2\gamma n) \cdot \left(1 - \frac{2}{1+\sqrt{1+4\gamma D}}\right)^{\ell}.$$

To analyze the running time, we note that in order to obtain a $(1 \pm \epsilon)$ multiplicative approximation to $p_v(G)$, it is sufficient to obtain a $\pm\epsilon/(1+\gamma n)$ additive approximation; this is because $p_v(G) \ge 1/(1+\gamma n))$ since the degree of each vertex in G is at most n. Now, as observed in Section 2.2, $p_v(G)$ always lies between the quantities $F_v(\mathbf{0}_\ell)$ and $F_v(\mathbf{1}_\ell)$, so in order to obtain a $\pm\epsilon/(1+\gamma n)$ approximation, it is sufficient to choose $\ell \ge a \log n$ large enough so that the right hand side of eq. (4.10) is at most $\epsilon/(1+\gamma n)$. Denoting by β the quantity in the parenthesis on the right hand side of eq. (4.10), we can ensure this by choosing

$$\ell \ge \frac{1}{\log(1/\beta)} \left[\log \frac{1+\gamma n}{\epsilon} + \log\log\left(\sqrt{1+2\gamma n}\right) + \frac{1}{\sqrt{1+4\gamma D}} \log c\right].$$

Further, given such an ℓ, the running time of the algorithm is $O(\sum_{i=1}^{\ell} N(v,\ell)) = O(\Delta^\ell)$, since this is the time it takes to expand the self-avoiding walk tree up to depth ℓ. Noting that $1/(\log(1/\beta)) = \sqrt{\gamma\Delta} + \Theta(1)$, we obtain an algorithm running in time

$$((1+\gamma n)/\epsilon)^{O(\sqrt{\gamma\Delta}\cdot\log\Delta)}$$

which provides a $(1 \pm \epsilon)$ multiplicative approximation for $p_v(G)$. The self-reducibility arguments referred to above (and provided in detail in Appendix A) show that this yields an algorithm for approximating the partition function up to a multiplicative factor of $(1 \pm \epsilon)$ with the same asymptotic exponent in the running time. This completes the proof. □

References

[1] Alm, S. E. Upper and lower bounds for the connective constants of self-avoiding walks on the Archimedean and Laves lattices. *J. Phys. A 38*, 10 (Mar. 2005), 2055–2080.

[2] Bandyopadhyay, A., and Gamarnik, D. Counting Without Sampling: Asymptotics of the Log-Partition Function for Certain Statistical Physics Models Random Structures and Algorithms. *Random Struct. Algorithms 33*, 4 (2008), 452–479.

[3] Bayati, M., Gamarnik, D., Katz, D., Nair, C., and Tetali, P. Simple deterministic approximation algorithms for counting matchings. In *Proc. 39th ACM Symp. Theory Comput.* (2007), ACM, p. 122–127.

[4] Duminil-Copin, H., and Smirnov, S. The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$. *Ann. Math. 175*, 3 (May 2012), 1653–1665.

[5] Dyer, M., and Greenhill, C. On Markov chains for independent sets. *J. Algorithms 35*, 1 (Apr. 2000), 17–49.

[6] Efthymiou, C. MCMC sampling colourings and independent sets of $\mathcal{G}(n, d/n)$ near uniqueness threshold. In *Proc. 25th ACM-SIAM Symp. Discret. Algorithms* (2014), SIAM, pp. 305–316. Full version available at arxiv:1304.6666.

[7] Galanis, A., Ge, Q., Štefankovič, D., Vigoda, E., and Yang, L. Improved inapproximability results for counting independent sets in the hard-core model. *Random Struct. Algorithms 45*, 1 (2014), 78–110.

[8] Gamarnik, D., and Katz, D. Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In *Proc. 18th ACM-SIAM Symp. Discret. Algorithms* (2007), SIAM, pp. 1245–1254.

[9] Godsil, C. D. Matchings and walks in graphs. *J. Graph Th. 5*, 3 (Sept. 1981), 285–297.

[10] Goldberg, L. A., Martin, R., and Paterson, M. Strong spatial mixing with fewer colors for lattice graphs. *SIAM J. Comput. 35*, 2 (2005), 486–517.

[11] Hammersley, J. M. Percolation processes II. The connective constant. *Math. Proc. Camb. Philos. Soc. 53*, 03 (1957), 642–645.

[12] Jerrum, M., and Sinclair, A. Approximating the permanent. *SIAM J. Comput. 18*, 6 (Dec. 1989), 1149–1178.

[13] Jerrum, M., Valiant, L. G., and Vazirani, V. V. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci. 43* (1986), 169–188.

[14] Kahn, J., and Kim, J. H. Random matchings in regular graphs. *Combinatorica 18*, 2 (Feb. 1998), 201–226.

[15] Li, L., Lu, P., and Yin, Y. Approximate counting via correlation decay in spin systems. In *Proc. 23rd ACM-SIAM Symp. Discret. Algorithms* (2012), SIAM, pp. 922–940.

[16] Li, L., Lu, P., and Yin, Y. Correlation decay up to uniqueness in spin systems. In *Proc. 24th ACM-SIAM Symp. Discret. Algorithms* (2013), SIAM, pp. 67–84.

[17] Madras, N., and Slade, G. *The Self-Avoiding Walk*. Birkhäuser, 1996.

[18] Martinelli, F., and Olivieri, E. Approach to equilibrium of Glauber dynamics in the one phase region I. The attractive case. *Comm. Math. Phys. 161*, 3 (Apr. 1994), 447–486.

[19] Martinelli, F., and Olivieri, E. Approach to

equilibrium of Glauber dynamics in the one phase region II. The general case. *Comm. Math. Phys. 161*, 3 (Apr. 1994), 487–514.

[20] Mossel, E., and Sly, A. Rapid mixing of Gibbs sampling on graphs that are sparse on average. *Random Struct. Algorithms 35*, 2 (2009), 250–270.

[21] Pönitz, A., and Tittmann, P. Improved upper bounds for self-avoiding walks in $\mathbb{Z}^d$. *Electron. J. Comb.*, vol. 7, Research Paper 21, 2000.

[22] Restrepo, R., Shin, J., Tetali, P., Vigoda, E., and Yang, L. Improved mixing condition on the grid for counting and sampling independent sets. *Probab. Theory Relat. Fields 156*, 1-2 (June 2013), 75–99.

[23] Sinclair, A., Srivastava, P., Štefankovič, D., and Yin, Y. Spatial mixing and the connective constant: Optimal bounds, Oct. 2014. Available at arxiv:1410.2595.

[24] Sinclair, A., Srivastava, P., and Thurley, M. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys. 155*, 4 (2014), 666–686.

[25] Sinclair, A., Srivastava, P., and Yin, Y. Spatial mixing and approximation algorithms for graphs with bounded connective constant. In *Proc. 54th IEEE Symp. Found. Comput. Sci.* (2013), IEEE Computer Society, pp. 300–309. Full version available at arxiv:1308.1762.

[26] Sly, A. Computational transition at the uniqueness threshold. In *Proc. 51st IEEE Symp. Found. Comput. Sci.* (2010), IEEE Computer Society, pp. 287–296.

[27] Sly, A., and Sun, N. The computational hardness of counting in two-spin models on d-regular graphs. In *Proc. 53rd IEEE Symp. Found. Comput. Sci.* (Oct 2012), pp. 361–369. Full version available at arxiv:1203.2602.

[28] Vera, J. C., Vigoda, E., and Yang, L. Improved bounds on the phase transition for the hard-core model in 2-dimensions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, vol. 8096 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013, pp. 699–713.

[29] Vigoda, E. A note on the Glauber dynamics for sampling independent sets. *Electron. J. Combin. 8*, 1 (Jan. 2001), R8.

[30] Weisstein, E. W. Self-avoiding walk connective constant. From *MathWorld* – A Wolfram Web Resource. http://mathworld.wolfram.com/Self-AvoidingWalkConnectiveConstant.html.

[31] Weitz, D. Counting independent sets up to the tree threshold. In *Proc. 38th ACM Symp. Theory Comput.* (2006), ACM, pp. 140–149.

A From probabilities to the partition function

In this section, we review some standard facts on how approximation algorithms for the marginal probabilities translate into approximation algorithms for the partition function (see, e.g, [8, 31]). We provide the calculations here for the case of the monomer-dimer model, and refer to Weitz [31] for similar calculations for the hard core model.

Let $v_1, v_2, \ldots, v_n$ be any arbitrary ordering of the vertices of G. Since the monomer-dimer partition function of the empty graph is 1, we then have

$$\begin{aligned} Z(G) &= \prod_{i=1}^{n} \frac{Z\left(G - \{v_1, \ldots, v_{i-1}\}\right)}{Z\left(G - \{v_1, \ldots, v_i\}\right)} \\ &= \prod_{i=1}^{n} \frac{1}{p_{v_i}\left(G - \{v_1, \ldots, v_{i-1}\}\right)}. \end{aligned} \tag{A.1}$$

Suppose, we have an FPTAS for the probabilities p_ρ which runs in time $t(n, 1/\epsilon)$ and produces an output $\hat{p}$ such that $p_\rho/(1+\epsilon) \le \hat{p} \le p_\rho$. Now, given $\epsilon \le 1$, we use the FPTAS in time $t\left(n, 2n/\epsilon\right)$ to compute an approximation $\hat{p}_i$ to the $p_{v_i}\left(G - \{v_1, \ldots, v_{i-1}\}\right)$. We then have for each i

$$\begin{aligned} \frac{1}{p_{v_i}\left(G - \{v_1, \ldots, v_{i-1}\}\right)} &\le \frac{1}{\hat{p}_i} \\ &\le \frac{1 + \epsilon/(2n)}{p_{v_i}\left(G - \{v_1, \ldots, v_{i-1}\}\right)}. \end{aligned}$$

By multiplying these estimates. we obtain an estimate $\hat{Z}$ of the partition function which satisfies

$$\begin{aligned} Z(G) \le \hat{Z} &\le Z(G)\left(1 + \frac{\epsilon}{2n}\right)^n \\ &\le Z(G) e^{\epsilon/2} \le Z(G)(1+\epsilon), \end{aligned}$$

where we use the condition $\epsilon \le 1$ in the last inequality. Thus, the total running time is $O\left(n \cdot t\left(n, 2n/\epsilon\right)\right)$, which is polynomial in n and $1/\epsilon$ whenever t is. Thus, it is sufficient to derive an FPTAS for the marginal probabilities in order to obtain an FPTAS for the partition function.

B Proofs omitted from Section 3

We reproduce here the proof of Lemma 3.2 from [25] for the convenience of the reader.

Proof of Lemma 3.2. Define $H(t) := f^{\phi}_{d,\lambda}(t\boldsymbol{x} + (1-t)\boldsymbol{y})$ for $t \in [0,1]$. By the scalar mean value theorem applied to H, we have

$$\begin{aligned} f^{\phi}_{d,\lambda}(\boldsymbol{x}) - f^{\phi}_{d,\lambda}(\boldsymbol{y}) &= H(1) - H(0) \\ &= H'(s), \text{ for some } s \in [0,1]. \end{aligned}$$

Let ψ denote the inverse of the message ϕ: the derivative of ψ is given by $\psi'(y) = \frac{1}{\Phi(\psi(y))}$, where Φ is the derivative of ϕ. We now define the vector $\boldsymbol{z}$

by setting $z_i = \psi(sx_i + (1-s)y_i)$ for $1 \le i \le d$. We then have

$$\begin{aligned}\left|f^{\phi}_{d,\lambda}(\boldsymbol{x}) - f^{\phi}_{d,\lambda}(\boldsymbol{y})\right| &= |H'(s)| \\ &= \left|\left\langle \nabla f^{\phi}_{\lambda,d}(s\boldsymbol{x} + (1-s)\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\right\rangle\right| \\ &= \Phi(f_{d,\lambda}(\boldsymbol{z}))\left|\sum_{i=1}^{d} \frac{x_i - y_i}{\Phi(z_i)} \frac{\partial f_{d,\lambda}}{\partial z_i}\right|, \\ &\qquad \text{using the chain rule} \\ &\le \Phi\left(f_{d,\lambda}(\boldsymbol{z})\right) \sum_{i=1}^{d} \frac{|y_i - x_i|}{\Phi(z_i)} \left|\frac{\partial f_{d,\lambda}}{\partial z_i}\right|.\end{aligned}$$

We recall that for simplicity, we are using here the somewhat non-standard notation $\frac{\partial f}{\partial z_i}$ for the value of the partial derivative $\frac{\partial f}{\partial R_i}$ at the point $\boldsymbol{R} = \boldsymbol{z}$. □

We now give the proof of the Lemma 3.4. The proof is syntactically identical to the proof of a similar lemma in [25], and the only difference (which is of course crucial for our purposes) is the use of the more specialized Lemma 3.3 in the inductive step. However, we reproduce the proof here for completeness.

Proof of Lemma 3.4. Recall that given a vertex v in $T_{\le C}$, T_v is the subtree rooted at v and containing all the descendants of v, and $F_v(\sigma)$ is the value computed by the recurrence at the root v of T_v under the initial condition σ restricted to T_v. We will denote by C_v the restriction of the cutset C to T_v.

By induction on the structure of T_ρ, we will now show that for any vertex v in T_ρ which is at a distance δ_v from ρ, and has arity d_v, one has

$$|\phi(F_v(\sigma)) - \phi(F_v(\tau))|^q \le M^q \sum_{u \in C_v} \alpha^{|u| - \delta_v}. \tag{B.2}$$

To see that this implies the claim of the lemma, we observe that since $F_\rho(\sigma)$ and $F_\rho(\tau)$ are in the interval $[0, b]$, we have $|F_v(\sigma) - F_v(\tau)| \le \frac{1}{L}|\phi(F_v(\sigma)) - \phi(F_v(\tau))|$. Hence, taking $v = \rho$ in eq. (B.2), the claim of the lemma follows from the above observation.

We now proceed to prove eq. (B.2). The base case of the induction consists of vertices v which are either of arity 0 or which are in C. In the first case (which includes the case where v is fixed by both the initial conditions to the same value), we clearly have $F_v(\sigma) = F_v(\tau)$, and hence the claim is trivially true. In the second case, we have $C_v = \{v\}$, and all the children of v must lie in C'. Thus, in this case, the claim is true by the definition of M.

We now proceed to the inductive case. Let $v_1, v_2, \ldots v_{d_v}$ be the children of v, which satisfy eq. (B.2) by induction. In the remainder of the proof, we suppress the dependence of ξ on ϕ and q. Applying Lemma 3.3 followed by the induction hypothesis, we then have, for some positive integer $k \le d_v$

$$\begin{aligned}&|\phi(R_v(\sigma)) - \phi(R_v(\tau))|^q \\ &\qquad \le \xi(k) \sum_{i=1}^{d_v} |\phi(R_{v_i}(\sigma)) - \phi(R_{v_i}(\tau))|^q, \\ &\qquad\qquad \text{using Lemma 3.3} \\ &\qquad \le M^q \xi(k) \sum_{i=1}^{d_v} \sum_{u \in C_{v_i}} \alpha^{|u| - \delta_{v_i}}, \\ &\qquad\qquad \text{using the induction hypothesis} \\ &\qquad \le M^q \sum_{u \in C_v} \alpha^{|u| - \delta_v},\end{aligned}$$

where in the last step we use $\xi(k) \le \alpha$ and $\delta_{v_i} = \delta_v + 1$. This completes the induction. □

C Symmetrizability of the message

In this section, we prove Lemma 4.1, which establishes the symmetrizability of the message ϕ defined in eq. (4.6). We begin with an auxiliary technical lemma.

Lemma C.1. *Let r and a satisfy $1 < r \le 2$ and $0 < a < 1$ respectively. Consider the functions $\gamma(x) := x^r(2-x)^r$ and $g(x) := \gamma(a - x) + \gamma(a + x)$. Note that g is even and is well defined in the interval $[-A, A]$, where $A := \min(a, 1 - a)$. Then all the maxima of the function g in the interval $[-A, A]$ lie in the set $\{-a, 0, a\}$.*

The lemma has the following simple consequence. Let $0 \le s_1, s_2 \le 1$ be such that $(s_1 + s_2)/2$ is constrained to be some fixed constant $a \le 1$. Then, applying the lemma with $s_1 = a - x, s_2 = a + x$, we see that $\gamma(s_1) + \gamma(s_2)$ is maximized either when $s_1 = s_2 = a$ or when one of them is 0 and the other is $2a$ (the second case can occur only when $a \le 1/2$).

Proof of Lemma C.1. Since g is even, we only need to analyze it in the interval $[0, A]$, and show that restricted to this interval, its maxima lie in $\{0, a\}$.

We begin with an analysis of the third derivative of γ, which is given by

$$\begin{aligned}\gamma'''(x) = -4r(r-1)(1-x)(1-(1-x)^2)^{r-2} \\ \cdot \left[\frac{3 - (2r-1)(1-x)^2}{1-(1-x)^2}\right].\end{aligned} \tag{C.3}$$

Our first claim is that γ''' is strictly increasing in the interval $[0, 1]$ when $1 < r \le 2$. In the case when $r = 2$, the last two factors in eq. (C.3) simplify to constants, so that $\gamma'''(x) = -12r(r-1)(1-x)$, which is clearly

strictly increasing. When $1 < r < 2$, the easiest way to prove the claim is to notice that each of the factors in the product on the right hand side of is a strictly increasing non-negative function of $y = 1 - x$ when $x \in [0, 1]$ (the fact that the second and third factors are increasing and non-negative requires the condition that $r < 2$). Thus, because of the negative sign, γ''' itself is a strictly *decreasing* function of y, and hence a strictly *increasing* function of x in that interval.

We can now analyze the behavior of g in the interval $[0, A]$. We first show that when $a > 1/2$, so that $A = 1-a \neq a$, g does not have a maximum at $x = A$ when restricted to $[0, A]$. We will achieve this by showing that when $1 > a > 1/2$, $g'(1-a) < 0$. To see this, we first compute $\gamma'(x) = 2rx^{r-1}(2-x)^{r-1}(1-x)$, and then observe that

$$\begin{aligned} g'(1-a) &= \gamma'(1) - \gamma'(2a-1) \\ &= -\gamma'(2a-1) < 0, \text{ since } 0 < 2a-1 < 1. \end{aligned}$$

We now start with the observation that $g'''(x) = \gamma'''(a+x) - \gamma'''(a-x)$, so that because of the strict monotonicity of γ''' in $[0, 1]$ (which contains the interval $[0, A]$), we have $g'''(x) > 0$ for $x \in (0, A]$. We note that this implies that $g''(x)$ is strictly increasing in the interval $[0, A]$. We also note that $g'(0) = 0$. We now consider two cases.

Case 1: $g''(0) \geq 0$ Using the fact that $g''(x)$ is strictly increasing in the interval $[0, A]$ we see that $g''(x)$ is also positive in the interval $(0, A]$ in this case. This, along with the fact that $g'(0) = 0$, implies that $g'(x) > 0$ for $x \in (0, A]$, so that g is strictly increasing in $[0, A]$ and hence is maximized only at $x = A$. As proved above, this implies that the maximum of g must be attained at $x = a$ (in other words, the case $g''(0) \geq 0$ cannot arise when $a > 1/2$ so that $A = 1 - a \neq a$).

Case 2: $g''(x) < 0$ Again, using the fact that $g''(x)$ is strictly increasing in $[0, A]$, we see that there is at most one zero c of g'' in $[0, A]$. If no such zero exists, then g'' is negative in $[0, A]$, so that g' is strictly decreasing in $[0, A]$. Since $g'(0) = 0$, this implies that g' is also negative in $(0, A)$ so that the unique maximum of g in $[0, A]$ is attained at $x = 0$.

Now suppose that g'' has a zero c in $(0, A]$. As before, we can conclude that g' is strictly negative in $[0, c]$, and strictly increasing in $[c, A]$. Thus, if $g'(A) < 0$, g' must be negative in all of $(0, A]$, so that g is again maximized at $x = 0$ as in Case 1. The only remaining case is when there exists a number $c_1 \in (c, A]$ such that g' is negative in $(0, c_1)$ and positive in $(c_1, A]$. In this case, we note that $g'(A) \geq 0$, so that—as observed above–we cannot have $A \neq a$. Further, the maximum of g in this case is at $x = 0$ if $g(0) > g(A)$, and at $x = A$ otherwise. Since we already argued that A must be equal to a in this case, this shows that the maxima of g in $[0, A]$ again lie in the set $\{0, a\}$. □

We now prove Lemma 4.1.

Proof of Lemma 4.1. We first verify the second condition in the definition of symmetrizability:

$$\lim_{p_i \to 0} \frac{1}{\Phi(p_i)} \left| \frac{\partial f_{d,\gamma}}{\partial p_i} \right| = \lim_{p_i \to 0} \frac{\gamma p_i (2 - p_i)}{\left(1 + \gamma \sum_{j=1}^{d} p_j\right)^2} = 0.$$

We now recall the program used in the definition of symmetrizability with respect to exponent r, with the definitions of Φ and $f_{d,\gamma}$ substituted:

$$\begin{aligned} \max \quad & \gamma^r f_{d,\gamma}(\boldsymbol{p})^{2r} \sum_{i=1}^{d} p_i^r (2 - p_i)^r, \qquad \text{where} \\ & \frac{1}{1 + \gamma \sum_{i=1}^{d} p_i} = B \\ & 0 \leq p_i \leq 1, \qquad 1 \leq i \leq d \end{aligned}$$

Since we are only interested in the values of $\boldsymbol{p}$ solving the program, we can simplify the program as follows:

$$\begin{aligned} \max \quad & \sum_{i=1}^{d} p_i^r (2 - p_i)^r, \qquad \text{where} \\ & \sum_{i=1}^{d} p_i = B' := \frac{1 - B}{\gamma B} \\ & 0 \leq p_i \leq 1, \qquad 1 \leq i \leq d \end{aligned}$$

We see that the feasible set is compact. Thus, if it is also non-empty, there is at least one (finite) optimal solution to the program. Let $\boldsymbol{y}$ be such a solution. Suppose without loss of generality that the first k co-ordinates of $\boldsymbol{y}$ are non-zero while the rest are 0. We claim that $y_i = y_j \neq 0$ for all $1 \leq i \leq j \leq k$.

For if not, let $i \neq j$ be such that $y_i y_j \neq 0$ and $y_i \neq y_j$. Let $y_i + y_j = 2a$. The discussion following Lemma C.1 implies that at least one of the following two operations, performed while keeping the sum $y_i + y_j$ fixed and ensuring that $y_i, y_j \in [0, 1]$ (so that all the constraints in the program are still satisfied), will increase the value of the sum $\gamma(y_i) + \gamma(y_j) = y_i^r (2 - y_i)^r + y_j^r (2 - y_j)^r$:

1. Making $y_i = y_j$, or
2. Making $y_i = 0$ (so that $y_j = 2a$). This case is possible only when $2a \leq 1$.

Thus, if $\boldsymbol{y}$ does not have all its non-zero entries equal, we can increase the value of the objective function while maintaining all the constraints. This contradicts the fact that $\boldsymbol{y}$ is a maximum, and completes the proof. □